Traduire en français le lexique du patrimoine artistique de la ville de Bologne : le sous-corpus comparable BER du projet LBC

Valeria Zotti

1. Introduction

Dans cette contribution, nous illustrerons une nouvelle phase du projet de recherche interuniversitaire *Lessico plurilingue dei Beni Culturali* (LBC) qui porte sur l'élargissement du corpus LBC Français[1] avec l'ajout d'un sous-corpus sur la ville de Bologne et sur la région Émilie-Romagne (Cetro, Zotti 2020, 88). Ce sous-corpus a été réalisé par une antenne de l'équipe LBC[2], l'unité de recherche LBC-CeSLiC de l'Université de Bologne[3], dans le cadre de deux projets de recherche financés : le premier, *UniCittà. Riscoprire Bologna attraverso l'Alma Mater* (UniCité. Redécouvrir Bologne à travers son Université), a été financé par la fondation Carisbo dans le cadre de l'appel à projets *Riscoprire la città*, au courant des années 2019-2021 ; le second, *UniVOCIttà: Voci digitali sull'unicità del patrimonio bolognese* (UniVOCité. Voix numériques sur l'unicité du patrimoine bolonais) vient d'être sélectionné par l'Université de Bologne parmi les projets de recherche explorant des thèmes de pertinence interdisciplinaire pour le Programme national de recherche italien (PNR).

[1] Consultable en libre accès sur : http://corpora.lessicobeniculturali.net/ (22/12/2022).

[2] Cf. le site web du projet LBC : https://www.lessicobeniculturali.net/fr/contenuti/l-équipe (22/12/2022).

[3] Née en mars 2018 au sein du CeSLiC (*Centro di Studi Linguistico-Culturali*) dans le but de formaliser le travail accompli au sein du projet LBC florentin.

Valeria Zotti, University of Bologna, Italy, valeria.zotti@unibo.it, 0000-0002-4920-4392

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Valeria Zotti, *Traduire en français le lexique du patrimoine artistique de la ville de Bologne : le sous-corpus comparable BER du projet LBC*, © Author(s), CC BY 4.0, DOI 10.36253/979-12-215-0061-5.14, in Valeria Zotti, Monica Turci (edited by), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali*, pp. 197-230, 2023, published by Firenze University Press, ISBN 979-12-215-0061-5, DOI 10.36253/979-12-215-0061-5

Dans la première partie de cette contribution, nous présenterons les finalités de ces deux projets, en nous arrêtant à la fois sur leurs spécificités et sur la composante de continuité avec les objectifs poursuivis par le projet LBC. Comme l'a affirmé Annick Farina dans une étude récente (Farina, Sini 2020, 77-78)[4], le projet LBC, né comme description plurilingue du patrimoine florentin, élargit désormais son horizon et s'ouvre à la description du patrimoine artistique italien, grâce au travail accompli, dans un premier temps, par l'équipe bolonaise.

Dans la seconde partie, nous présenterons le produit de ce travail, à savoir le sous-corpus Bologne et Émilie-Romagne (dorénavant BER) en son état actuel. Nous illustrerons son contenu tant sur le plan quantitatif (nombre de textes et de mots) que sur le plan qualitatif (nature des textes, sujets et aires géographiques couverts, données lexico-terminologiques), en nous appuyant sur la méthodologie offerte par la linguistique de corpus. Nous essayerons de montrer, dans une perspective plus large, quel sera l'apport de ce sous-corpus pour compléter la description de la terminologie artistique donnée par le corpus monolingue LBC Français.

Dans la troisième partie, nous nous pencherons sur les applications et les résultats fournis par la méthode d'analyse choisie. Nous montrerons dans quelle mesure l'approche adoptée (*corpus-driven*) permet de « faire émerger de manière inductive des savoirs linguistiques » (Williams 2005: 13), concernant notamment le phénomène de la synonymie (diastratique et diatopique) en terminologie (*arcades / portiques / galerie / loge*), ainsi que de suggérer de nouvelles stratégies pour la traduction de termes désignant des réalités culturo-spécifiques (*portiques*) qui ne sont pas prises en compte dans les ressources plurilingues actuellement disponibles. Cette réflexion sur l'apport des corpus pour la traduction de la terminologie artistique, qui ne s'appuie ici que sur l'exploration des corpus monolingues comparables LBC, sera complétée, dans la prochaine étape de notre recherche, par l'analyse des bases parallèles de traduction de l'italien vers le français et vers l'anglais des *Vies* de Vasari[5], que nous évoquerons à la fin.

2. Le projet LBC et les projets dérivés UniCittà et UniVOCIttà

Le projet de recherche *Lessico plurilingue dei Beni Culturali* est né en 2013 au sein du département de Langues, Littérature et d'Études Interculturelles de l'Université de Florence, pour répondre à l'urgence, perçue par Annick Farina et Marcello Garzaniti, de fournir aux professionnels de l'art des ressources fiables pour la description plurilingue du riche patrimoine artistique florentin et toscan[6].

[4] «Questi dizionari, come i corpora che li alimentano [...] nascono per consentire sia a un pubblico specializzato che a un pubblico più ampio di svolgere ricerche, qualitative e quantitative, sul patrimonio artistico delle città di Firenze e della Toscana e, in una seconda fase, del patrimonio artistico italiano».

[5] Travail en cours sous la supervision de Daniel Henkel et de l'auteure de la présente contribution.

[6] Pour des informations détaillées et une description complète du projet LBC, voir la page web consacrée sur http://www.lessicobeniculturali.net/contenuti/il-progetto/818 (22/12/2022), ainsi que les nombreuses publications dans la collection « Lessico Multilingue dei Beni

À partir de ce projet interuniversitaire, dans lequel plusieurs chercheurs de l'Université de Bologne se sont engagés dès son origine, deux projets satellites se sont développés : le premier, *UniCittà. Riscoprire Bologna attraverso l'Alma Mater*[7], financé entre 2019 et 2021 par la Fondation *Cassa di Risparmio di Bologna*, porte sur la valorisation du patrimoine matériel et immatériel de l'Alma Mater de Bologne, l'Université la plus ancienne d'Europe ; le second, *UniVOCIttà: Voci digitali sull'unicità del patrimonio bolognese*, qui en est la continuation naturelle, vient de recevoir un financement pour une durée de 18 mois (juillet 2022-décembre 2023) par l'Université de Bologne dans le cadre du programme de financements AlmaIdea, et a pour but de s'ouvrir à la description du patrimoine de toute la ville de Bologne et de sa région.

Dans les deux projets bolonais, sous l'égide du projet LBC, le plan de travail et la méthodologie adoptée suivent en grande partie ceux du projet LBC en ce qui concerne la collecte et la numérisation de textes pour les trois typologies textuelles identifiées par les concepteurs du projet (Billero 2020 ; Farina, Nicolas Martinez 2020). Les finalités coïncident aussi, étant donné que ces deux projets dérivés visent la valorisation du patrimoine local. Dans le premier cas (*UniCittà*), on s'intéresse à l'immense capital artistique immobilier (bibliothèques, églises, musées universitaires, instituts religieux, jardin botanique, etc.) et mobilier (collections muséales et libraires, archives historiques, etc.), accumulé par l'Université de Bologne au cours des siècles depuis sa fondation en 1088 (Perotto, Zotti 2022). Dans le second cas (*UniVOCIttà*), c'est la richesse artistique et culturelle de la ville de Bologne, composée de monuments, musées et palais historiques, moins célèbres dans les circuits touristiques traditionnels, qui est mise en valeur, ainsi que le précieux patrimoine naturel de ses alentours et de toute la région qui, quant à lui, mérite d'être préservé[8].

Ces deux projets ne se distinguent de LBC que sur le plan des produits finaux envisagés. Ils répondent à l'objectif, sollicité par les organismes de financement, de 'troisième mission' de l'Université, communément appelé « service à la société ». C'est la raison pour laquelle ces deux produits s'adressent surtout au grand public et à l'industrie du tourisme et moins à un public spécialisé de linguistes.

Le produit final du projet *UniCittà* est la base de données textuelles homonyme *Corpus UniCittà*[9], qui permet d'obtenir des fragments de textes de longueur variable en quatre langues (anglais, espagnol, français et russe), décrivant le patrimoine de l'Université du point de vue des voyageurs étrangers du passé. Ces fragments ont été étiquetés avec des mots-clés correspondant à des catégories thématiques à l'aide de Atlas.ti, un logiciel performant pour l'analyse

Culturali » chez l'éditeur Firenze University Press : https://books.fupress.com/scientific-board/lessico-multilingue-dei-beni-culturali/67 (22/12/2022).

7 Voir le site web du projet UniCittà: https://site.unibo.it/unicitta/it (22/12/2022).

8 Nous pensons par exemple aux chemins reliant Bologne et Florence sur les Apennins, comme le célèbre *Chemin des Dieux*.

9 Disponible sur https://corpusunicitta.it/ (22/12/2022).

qualitative des données textuelles, comme l'a observé Williams (2020 : 200)[10]. L'objectif, ici, est de fournir du matériel documentaire aux opérateurs touristiques, en particulier aux commissaires d'expositions, et aux organisateurs de visites guidées[11]. Ce matériel offrirait aux visiteurs d'aujourd'hui le regard des illustres visiteurs étrangers du passé (souvent des hommes et femmes de lettres et de sciences venus étudier ou enseigner à l'Université de Bologne) sur le patrimoine universitaire[12].

Le projet *UniVOCIttà*, en cours, se propose d'élargir le champ d'application du projet précédent, en recueillant les "voix" qui attestent de l'unicité du patrimoine de Bologne et de l'Émilie-Romagne, en termes d'exceptionnalité et d'altérité par rapport à un regard étranger[13]. Ce projet répond à différents défis posés par le PNR 2021-2027 : l'objectif est de promouvoir une synergie entre les compétences scientifiques, l'innovation technologique et le système productif, en l'occurrence le secteur du tourisme culturel, en impliquant les régions, ainsi que de développer des technologies pour soutenir le patrimoine culturel moins connu et de participer à sa conservation et valorisation par la numérisation afin d'en favoriser une large accessibilité.

Dans le cadre de ce projet, l'unité de recherche LBC-CeSLiC du Département de Langues, Littératures et Cultures Étrangères de l'Université de Bologne, en collaboration avec le Département d'Informatique de la même Université, envisage la création d'un site web mobile réactif, pouvant favoriser la promotion du tourisme patrimonial de la ville de Bologne et de sa région. Ce système, développé par un ingénieur informatique qui bénéficiera d'une allocation de recherche d'une durée de 12 mois (début : 1er janvier 2023), sous la supervision du professeur Fabio Vitali, expert dans la conception de vocabulaires sémantiques pour l'annotation de documents textuels et d'œuvres littéraires, sera en mesure de générer des visualisations dynamiques des fragments du corpus textuel, mettant en évidence les interconnexions linguistiques, culturelles et sémantiques. Les résultats seront présentés sous forme de cartes numériques géo-référencées qui pourront être facilement explorées et analysées par différents types d'utilisateurs, dont les visiteurs.

10 « Atlas.ti provides a useful means of analysing a text and extracting key information » (Williams 2020 : 200).

11 Comme celle proposée par le *Museo della Storia di Bologna* qui porte sur l'importance de l'Université dans l'économie de la ville. Cf. site web du musée : https://www.bolognawelcome.com/it/luoghi/musei-e-gallerie-darte/palazzo-pepoli-museo-della-storia-di-bologna (22/12/2022).

12 L'exemple souvent donné pendant les activités de dissémination dans les écoles et dans les bibliothèques publiques de la ville qui ont été déjà réalisées au cours de la période 2019-2021 est celui des deux tours de Bologne. Vues des hauteurs voisines les deux tours sont tantôt aimées et comparées aux mâts d'un vaisseau immense, par Joseph Louis Théodore Moyne, tantôt détestées comme par Charles Dickens.

13 À titre d'exemple, les descriptions suggestives de la perception des portiques de Bologne par d'autres regards étrangers apportent un éclairage inédit pour la promotion de ce patrimoine de l'Unesco (cf. A. Robida, *Les vieilles villes d'Italie: notes et souvenirs*, 1878).

Bien que l'activité en aval de création de ces deux produits se distancie forcément de la méthodologie de LBC[14] pour ce qui relève des dernières phases (étiquetage sémantique, géolocalisation, modalité de navigation, types d'utilisateurs, etc.), tout le travail en amont de collecte, numérisation et catégorisation des textes correspond exactement au projet LBC. Il est donc prévu que le projet UniVOCItà s'adresse aussi à la communauté scientifique de linguistes, et que cette expérience d'humanités numériques alimente l'élargissement du corpus LBC et converge dans la création d'un sous-corpus textuel plurilingue sur Bologne et l'Émilie-Romagne prochainement intégré dans le corpus LBC Français. Un grand nombre de textes a été déjà collecté dans le cadre du premier projet sur l'Université de Bologne au cours des années 2019-2021 et l'objectif d'UniVOCItà est d'élargir le corpus pour chaque langue concernée.

Dans la prochaine section, nous présenterons le sous-corpus Bologne et Émilie-Romagne en son état actuel. Nous illustrerons les critères adoptés pour sa constitution et son contenu, sur le plan quantitatif et qualitatif, en nous appuyant sur les données extraites au moyen de l'outil d'exploration de corpus textuels *Sketch Engine*. Nous évoquerons aussi quel sera l'apport de ce sous-corpus pour l'élargissement et le rééquilibrage futurs du corpus monolingue LBC Français.

3. Présentation du sous-corpus BER et complémentarité avec le corpus LBC

Le sous-corpus BER a été réalisé sous notre supervision par l'équipe LBC-CeSLiC, composée de cinq chercheurs travaillant sur quatre langues étrangères[15], avec la collaboration de plusieurs étudiants-stagiaires[16] qui se sont occupés, dans une première phase, de la recherche des ouvrages, de leur conversion en version électronique et de leur mise au propre pour rendre les documents analysables par les logiciels de TAL utilisés. Un travail de lecture important a été ensuite effectué par les stagiaires pour l'analyse du contenu des textes et leur étiquetage selon des critères sémantiques à l'aide du logiciel Atlas.ti. Bien que l'étiquetage ait été orienté vers la constitution de la base *UniCittà*, cette lecture approfondie s'est avérée très utile aussi pour la catégorisation des différents types de textes. Comme pour tous les corpus comparables LBC, les textes ont été choisis en fonction de leur importance historico-culturelle, de leur diffusion au niveau international et du prestige accordé au patrimoine culturel italien (cf. Billero, Martinez 2018 : 208). L'équipe a donc privilégié la qualité des données plutôt que leur quantité, étant donné que,

[14] De fait, des projets annexes ont aussi été entamés par l'équipe LBC de l'Université de Florence en utilisant les ressources LBC, comme le travail en collaboration avec la bibliothèque de Sciences Sociales de l'Université de Florence, en cours, pour créer ou compléter les fiches Wikidata relatives œuvres d'art (français/italien) et mettre en place leur géolocalisation, en utilisant les traductions de Vasari comme source de référence.

[15] Ana Pano Alamán pour l'espagnol, Monica Turci et Antonella Luporini pour l'anglais, Monica Perotto pour le russe et Valeria Zotti pour le français.

[16] Pour la langue française : Alice Curti, Francesca Faccioli, Aurora Ghelli, Valentina Toia, Matilde Signorelli, Carlo Garavaglia et Martina Izzi.

contrairement aux corpus de langue générale, la dimension d'un corpus comparable spécialisé n'est pas proportionnelle à sa qualité (Friginal, Hardy 2014).

3.1. Critères de constitution et données quantitatives et qualitatives

La recherche des textes a reposé principalement sur la consultation d'une anthologie italienne qui recense minutieusement les références bibliographiques des récits et témoignages de voyageurs en Émilie et en Romagne au XVIIIe siècle : *Viaggi e viaggiatori del Settecento in Emilia e in Romagna* (Cusatelli 1986). La plupart des textes libres de droits d'auteur ont été ensuite repérés dans les principales bibliothèques numériques en accès libre (*Gallica, Wikisource*), alors que les textes les plus récents ont été repris des plateformes de diffusion de publications scientifiques (*Persée, Érudit*), de sites web et blogues de tourisme et de la presse en ligne. Dans quelques cas, des œuvres disponibles en format papier ont été numérisées avec un convertisseur OCR et post-traitées manuellement pour corriger les éventuelles erreurs générées par ce système[17]. Tous les textes du corpus ont été ainsi convertis au format .doc et .txt pour qu'ils soient analysables, à ce stade, sur *Sketch Engine*, en vue de leur future intégration dans *NoSketchEngine*, le logiciel de gestion des corpus utilisé pour le projet LBC (cf. Billero 2020 : 29-30).

Le sous-corpus BER Français, qui a été collecté au cours du projet *UniCittà*, afin d'extraire les fragments concernant la description du patrimoine de l'Université de Bologne consultables dans la base de données <corpusunicittà.it>, est constitué à l'heure actuelle (juillet 2022) de 201 textes, tous rédigés en français langue originale (aucune traduction), pour un total de plus de 600 000 *tokens*, répartis de la manière suivante pour chaque typologie textuelle (Fig. 1) :

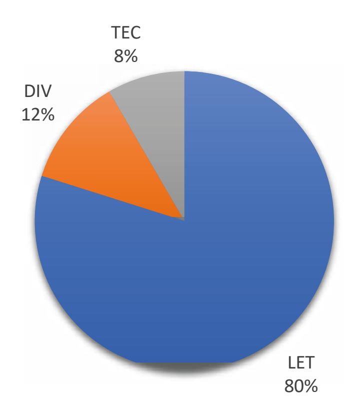


Fig. 1 Pourcentage de *tokens* par typologie textuelle dans le corpus BER Fr.

[17] Le post-traitement par reconnaissance optique de caractères (OCR) implique des étapes de nettoyage des données pour les documents qui ont été numérisés. Une étape de ce processus est l'identification et la correction des fautes d'orthographe et de grammaire générées en raison des défauts du système OCR.

Tab. 1. Données quantitatives du corpus BER Fr par rapport au Corpus LBC Fr.

Corpus BER	n. doc. BER	tokens BER	words BER	% words BER	words LBC	% words LBC	BER > LBC FR
DIV	69	71 269	60 887	*11,74 %*	346 148	10,92 %	+ 1,92 %
LET	98	483 587	413 856	*79,86 %*	1 847 208	58,35 %	+ 13,07 %
TEC	34	49 870	43 475	*8,38 %*	259 009	8,42 %	+ 1,37 %
DIZ	-	-	-	-	704 622	22,26 %	-
Total	*201*	*604 726*	*518 218*	*100 %*	*3 164 995*	*100 %*	*+ 16,36 %*

Le tableau 1 montre que le pourcentage de *words* (c'est-à-dire des *tokens* qui commencent par une lettre de l'alphabet sur SketchEngine) de chaque catégorie textuelle sur la totalité du sous-corpus BER correspond au pourcentage de *words* pour la même catégorie dans le corpus LBC (ex. 11,74% DIV dans BER = 10,92% DIV dans LBC FR). L'accroissement dû au sous-corpus BER au sein du corpus LBC Français est, à l'état actuel, de 16% de *words* au total (cf. la dernière colonne du tableau 1). L'objectif est d'élargir BER dans le cadre du nouveau projet *UniVOCIttà* et d'atteindre le chiffre d'au moins 800 000 *tokens* d'ici la fin du mois de décembre 2022, ce qui correspondra finalement à un accroissement du corpus LBC Français d'environ 20% pour chacune des 4 langues concernées dans le corpus BER, outre un grand nombre de textes supplémentaires préparés par d'autres équipes et également en attente d'être intégrés[18]. Ce sous-corpus pourra ainsi être intégré pendant l'année 2023 au corpus Français LBC, qui compte à l'heure actuelle 3 164 995 *words* et 3 818 747 *tokens*, dont environ 10% est constitué par des traductions de l'italien au français des *Vite* de Vasari (Farina, Sini 2020: 81).

La méthode adoptée pour la sélection des données textuelles correspond exactement à celle du projet LBC et se fonde sur la délimitation des typologies textuelles, des critères chronologiques et d'une zone géographique (cf. Farina, Nicolás Martínez 2020: 13).

3.1.1. Typologies textuelles

Les textes ont été recueillis en adoptant les catégories établies pour le projet LBC, à savoir : textes de vulgarisation (DIV, abréviation du mot italien *divulgativo*), comprenant les guides et les blogues touristiques ; textes techniques adressés à des spécialistes (TEC), entre autres les textes sur les techniques artistiques

[18] Les membres de l'équipe LBC-CeSLiC de l'Université de Bologne se sont donné cette limite commune, car pour certaines langues, par ex. le russe, il est moins aisé de repérer suffisamment de textes de voyageurs étrangers ayant visité la ville de Bologne et la région Émilie-Romagne dans le passé. En effet, comme dans le projet général LBC, la quantité de documents recensés dans les différentes langues dépend tant de l'accessibilité aux textes que de l'intérêt envers le patrimoine artistique italien qui a varié au cours des siècles et selon les pays d'origine des différents auteurs.

et les textes de critique d'art ; et textes littéraires (LET), catégorie qui englobe en premier lieu les récits de voyage mais aussi des correspondances et des textes de fiction[19]. La catégorie « dictionnaires » (DIZ) fait défaut dans le sous-corpus BER car, dans les faits, il s'agit d'une catégorie transversale ne se rapportant pas à une aire géographique donnée et recouvrant différents domaines reliés aux Beaux-Arts (architecture, peinture, sculpture).

Le sous-corpus DIV-BER comprend des textes tirés surtout de guides touristiques en ligne (ex. *Orangesmile*), de sites web de tourisme, de blogues de voyageurs (*Yummyplanet*), de photographes, mais aussi de sociologues[20], de sites web de musées et de revues de vulgarisation et culturelles (*Télérama*). Plusieurs textes proviennent du blogue français sur le tourisme en Italie, *Bellitalie.org*. Ils décrivent les monuments principaux de la ville de Bologne, ses églises (Basilique Saint François, Basilique Saint-Jacques Majeur, Basilique de Saint Stéphane ou des « sept églises », etc.), ses musées (ex. Musée archéologique de Bologne), les lieux les plus connus de son Université (l'Archiginnasio, palazzo Poggi, via Zamboni, les musées universitaires), ainsi que les peintres contemporains nés dans la ville (Giorgio Morandi). Le regard des visiteurs québécois sur la ville de Bologne est aussi présent à travers les textes transcrits des chroniques de *RadioCanada* et des articles de divers titres de la presse québécoise (*Le Clairon*, *Le Devoir*, *Le Soleil*, *La Presse*, etc.) : 21 textes de la catégorie DIV ont été en fait puisés dans la Bibliothèque et Archives numériques du Québec (BAnQ), grâce à la collaboration de Myriam Vien et de quelques stagiaires du CISQ (Centre InterUniversitaire des Etudes Québécoises), dont le siège central est situé à l'Université de Bologne.

Le sous-corpus LET-BER comprend principalement des journaux et carnets de voyages, des correspondances[21] et des extraits de romans. Les auteurs sont des écrivains et philosophes (à titre d'exemple, René Bazin, Alfred Driou, Antoine Claude Pasquin Valéry, Hippolyte Taine), des artistes (Charles-Nicolas Cochin, peintre et dessinateur), des intellectuels et personnalités politiques (Charles de Rémusat, Honoré Beaugrand), des scientifiques (Charles-Marie de La Condamine et Jérôme de La Lande, astronomes), des historiens de l'art (Paul Marmottan). Les femmes de lettres sont bien représentées (Noémie Dondel du Faouëdic, Anne-Marie du Boccage, Madame de Genlis, Elisabeth De

[19] Il faut tenir compte du fait que, pour certains textes, l'attribution à l'une des catégories établies n'est pas simple ni univoque, car certains textes peuvent faire partie de plusieurs catégories ou de catégories différentes d'une langue à l'autre. En effet, au fil du temps, le pacte de lecture peut changer et les textes rédigés avec une finalité déterminée peuvent rentrer aujourd'hui dans une catégorie différente (cf. Farina 2020: 91).

[20] Plusieurs textes sont tirés du blogue très médiatisé *Histoires d'Universités* créé par le sociologue Pierre Dubois : http://blog.educpros.fr/pierredubois (22/12/2022).

[21] «Anche la classificazione di testi epistolari nella categoria "Letterario-Biografico" è stata oggetto di discussione: le prime corrispondenze introdotte nei corpora sono state quelle di viaggiatori del *Grand Tour* francese in uno stile molto simile a quello dei quaderni di viaggio, si è quindi pensato di inserirli nella stessa categoria» (Farina, Nicolás Martínez 2020, 11).

Gonzague, Madame de Reck), et, parmi elles, notons la présence remarquable de la célèbre écrivaine Madame De Staël (1807) qui a consacré une partie de son roman *Corinne ou L'Italie* au récit de la visite de Bologne des deux protagonistes. Un chapitre de l'ouvrage de Stendhal *Écoles italiennes de peinture* (1932), paru à titre posthume, est consacré à l'« École de Bologne » et a été classé, pour l'instant, parmi les textes littéraires pour des raisons purement biographiques. Il sera probablement déplacé dans la catégorie des textes techniques à la suite d'un travail de révision de la classification des textes qui sera effectué sous peu avec la responsable du Corpus LBC Français, Annick Farina (cf. Farina, Nicolás Martínez 2020: 11)[22]. Les écrits du peintre, illustrateur et critique d'art Jean-Joseph Taillasson (1807) sur les peintres bolonais Annibal Carrache et Le Guerchin seront aussi sans doute placés dans la catégorie TEC, en raison du caractère technique du langage employé.

Outre les descriptions concernant l'architecture de la ville de Bologne, telles que les arcades, les portes, les tours, les riches demeures, les jardins intérieurs, les théâtres et les palais, on trouve dans cette catégorie, à la différence du sous-corpus DIV, nombre de fragments textuels consacrés aux lieux de l'Université[23], où ces voyageurs savants ont parfois assisté à des cours ou enseigné, ainsi que leurs réflexions sur la position prééminente de cette Université, sur la qualité et la variété de l'enseignement, sur l'esprit libéral qui, avec la beauté de la ville, attirent un grand nombre d'étudiants de toutes les époques. Des descriptions poétiques alternent ainsi avec des comptes rendus pointus et détaillés du patrimoine de la ville de Bologne et de ses alentours[24] ; on y trouve aussi de véritables critiques des œuvres d'art, comme celles de Charles-Nicolas Cochin (1758) concernant les peintres de l'école bolonaise. Les artistes les plus évoqués sont les artistes du XVIe siècle, comme Nicolò dell'Abate, qui a réalisé les fresques du palais de l'Institut des Sciences (aujourd'hui *palazzo Poggi*), et Pellegrino Tibaldi qui a exécuté ses plafonds décorés. Les témoignages récoltés dans ces textes littéraires dévoilent aussi la stupeur et l'admiration ressenties par les voyageurs francophones quant à la position prééminente des femmes enseignant à l'Université de Bologne dès le XIVe siècle (Novella d'Andrea) jusqu'au XVIIIe (Maria Agnesi, Laura Bassi, Clotilde Tambroni), ce qui fait de ce sous-corpus aussi

[22] «Da queste osservazioni deriva una eterogeneità fra corpora che vorremmo limitare negli sviluppi futuri del progetto. Infatti, l'analisi della distribuzione delle tipologie di testi scelti in ogni corpus e dei secoli rappresentati alla fine di questa prima fase di costituzione dei corpora potrà permettere una più ampia omogeneizzazione in futuro, consentendo lavori di comparazione dei testi» (Farina, Nicolás Martínez 2020, 11).

[23] À titre d'exemple : la Bibliothèque, les cliniques médicale et chirurgicale, l'amphithéâtre anatomique, les cabinets d'anatomie et d'histoire naturelle, la collection pathologique, les musées obstétrique, zoologique, minéralogique, l'observatoire, le jardin botanique, le collège des Flamands et le collège espagnol.

[24] Nous avons aussi intégré dans cette catégorie un chapitre d'un essai écrit par Bertrand Jestaz (1966), qui analyse les voyages en Italie de Robert de Cotte, un architecte du XVIIe siècle, et présente une description très technique de quelques œuvres architecturales bolonaises, notamment des églises S. Salvatore, S. Paolo et de l'église métropolitaine S. Pietro.

un dépôt de connaissances sur l'histoire de la culture et de la société italiennes d'un point de vue étranger[25].

Le sous-corpus TEC comprend des textes hétérogènes : catalogues, essais, articles de revues spécialisées, principalement dans les domaines de l'architecture (*L'architecture d'aujourd'hui*), de la peinture et de la gravure (*Nouvelles de l'Estampe*)[26]. Les articles scientifiques, qui couvrent surtout les domaines de l'architecture et de l'ingénierie du bâtiment, décrivent minutieusement les parties constitutives de certains établissements bolonais, parmi lesquels le *Teatro Comunale* de Bologne réalisé par l'architecte Antonio Galli, et fournissent des exemples des différents styles architecturaux de la ville (roman, baroque, fasciste, etc.). D'autres articles, tirés de revues spécialisées dans les domaines touchant aux Musées et aux Beaux-Arts (*Bulletin des musées de France, Gazette des Beaux-Arts*), décrivent quelques œuvres des peintres de l'école bolonaise, comme celles de Nicolò dell'Abate, déjà mentionné, et des disciples des Carrache. Dans ce sous-corpus sont aussi présents des textes tirés de revues d'histoire moderne et contemporaine, traitant de différents sujets relatifs aux « arts mécaniques » : les corporations des peintres, la censure des images, le marché de l'art à Bologne à différentes époques. Des sites web spécialisés, comme celui du laboratoire *Studio Leonardo*, couvrent aussi le domaine de la restauration des biens artistiques. Pour finir, quelques articles tirés de L'*Encyclopédie Larousse* ont été intégrés, comme celui sur l'« Académie des Carrache », en vue d'en incorporer d'autres d'ici la fin du projet UniVOCIttà.

3.1.2. Critères chronologiques

Afin de pouvoir être intégrés dans les corpus LBC, les textes doivent respecter des critères chronologiques et donc dater d'une période historique donnée qui couvre six siècles : de la Renaissance (XVIe siècle) à nos jours (XXIe siècle). Comme l'a observé Farina (2020 : 83), à l'heure actuelle, les siècles les plus représentés dans le corpus LBC Français sont le XIXe et le XXIe. La prédominance de ces deux siècles est due à des raisons d'accessibilité des textes : la large présence de la littérature de voyage du XIXe siècle et surtout de l'époque romantique dépend, au-delà de la vaste production de cette époque, d'une plus grande disponibilité de textes numérisés qui sont désormais libres de droits. De même, pour l'époque contemporaine une grande quantité de textes peut être repérée en ligne et téléchargée.

Dans le sous-corpus BER, la distribution des textes entre les siècles n'est pas homogène non plus. Actuellement, les textes recueillis couvrent un cadre temporel qui va du XVIIIe au XXIe siècle, distribués de la sorte pour chaque catégorie (Tab. 2) :

[25] Ce potentiel a été justement exploité dans le projet UniCittà déjà illustré.

[26] À titre d'exemple, un texte présente la technique de la gravure, développée à Bologne par les Carrache, qui ont été des maîtres aussi bien en peinture qu'en gravure. L'auteur de ce texte (Huchet 2011) décrit avec admiration les œuvres des Carrache, citant aussi les palais de Bologne où elles se trouvent.

Tab. 2. Distribution des textes par siècle et par catégorie dans le sous-corpus BER.

	DIV	TEC	LET
BER Français	-	XVIII[e] : 1	XVIII[e] : 55
	XIX[e] : 2	XIX[e] : 5	XIX[e] : 29
	XX[e] : 22	XX[e] :14	XX[e] : 13
	XXI[e] : 45	XXI[e] : 14	XXI[e] : 1
Tot.	69	34	98

Dans la catégorie DIV-BER, la plupart des documents (45) sont tirés de sites-web de voyage, et dans quelques cas, de revues généralistes en ligne, du début du XXI[e] siècle, surtout des années 2019-2021 qui correspondent à la période de déroulement du premier projet bolonais. Il est intéressant de remarquer que les textes du XX[e] siècle (22) viennent en majorité (18) de la presse québécoise de l'époque (cf. plus haut) et portent pour la plupart sur l'intérêt de la ville de Bologne en tant que modèle d'urbanisme et d'exemple de rénovation urbaine. Le XIX[e] siècle n'est représenté pour l'instant que par deux articles tirés encore de la presse québécoise (*La Gazette de Joliette*, 1870 ; *La Justice*, 1890), ce qui permet d'inférer que le regard francophone nord-américain sur le patrimoine bolonais est bien représenté dans le sous-corpus BER.

En ce qui concerne la catégorie TEC-BER, alors que les siècles les plus couverts dans le Corpus LBC Français dans la même catégorie sont le XVII[e] (80%), suivi du XIX[e] (13.3%) et du XX[e] (6,7%)[27], dans le sous-corpus BER les 34 textes présents sont pour l'instant répartis sur deux siècles (14 pour le XX[e] et 14 pour le XXI[e]). Un seul texte a été rédigé au XVIII[e] siècle, alors que le XVI[e] n'est pas encore représenté, comme pour le Corpus LBC Français (cf. Tab. 2), bien que des textes du XVIe siècle soient en cours d'intégration, comme un court texte sur Bologne tiré du carnet de voyages de Michel de Montaigne.

Dans la catégorie LET-BER, la grande majorité des textes (55) date du XVIII[e] siècle. Cela est dû au fait que, comme nous l'avons illustré plus haut, les références bibliographiques proviennent d'une anthologie de récits de voyageurs francophones de ce siècle (Cusatelli 1986). C'est cette caractéristique en particulier qui rend ce sous-corpus complémentaire au corpus LBC Français, étant donné que ce siècle y est très peu représenté (seulement 3,6%), alors que le XIX[e] siècle y prime avec 61,6% des documents.

Comme l'a observé Farina (2020: 90), la recherche d'un équilibre formel des données du corpus constitue un défi. Nous rappelons à ce propos que, pour la représentativité et l'équilibre d'un corpus comparable, il est nécessaire d'intégrer plusieurs typologies textuelles, genres, registres et auteurs de la langue dont le corpus se veut représentatif, et que les échantillons de chaque typologie doivent avoir une dimension identique ou le plus comparable possible (McEne-

[27] Données reprises du tableau 4 dans Farina (2020, 85).

ry, Hardy 2012). Le corpus comparable LBC français est en cours de construction et ses premières applications ont mis en lumière quelques limites relatives à cet aspect. Son élargissement avec ce sous-corpus vise à combler l'écart existant entre les typologies textuelles et la différente représentation des siècles et à réduire les sur-représentations dans le corpus, comme celle du *Dictionnaire raisonné de l'architecture française* de Viollet Le Duc (1854-1868) mentionnée par Farina (2020 : 80). Un plus grand équilibre garantira des résultats plus fiables pour mener des études contrastives dans le corpus plurilingue LBC[28].

3.1.3. Zone géographique

Alors que la version 1.0 du Corpus LBC Français est caractérisée par la priorité accordée au lexique de l'art et du patrimoine artistique florentin et toscan en français, la portée géographique de ce corpus spécialisé changera à l'avenir pour couvrir toute l'Italie et d'autres cultures (cf. Farina 2020: 91, Billero 2020: 21)[29]. Le sous-corpus BER se veut représentatif des discours sur le patrimoine artistique et culturel de Bologne et de sa région d'appartenance, ce pourquoi le critère géographique adopté pour la sélection des textes prévoit la présence du nom de la ville de Bologne et des villes de l'Émilie-Romagne, des principaux fleuves et rivières qui traversent la région (Savena, Panaro, Reno, Samoggia), des chaînes de montagnes (Apennins Tosco-Émiliens), ainsi que des artistes nés dans cette région. Nous rapportons ici, par ordre de fréquence, la liste des toponymes les plus présents dans le corpus BER dans la langue où elles figurent dans les textes (italien ou français) avec, en italique, leurs variantes :

Tab. 3. Fréquence des toponymes dans les trois catégories du sous-corpus BER.

Item	***DIV***	***LET***	***TEC***
Bologne	634	914	118
Bologna	32	10	2
Felsina	4	7	1
Bononia	2	6	-
Bononiae	2	-	2
Parme	11	339	-
Parma	2	57	-
Parmes	-	1	-

[28] Composé à l'heure actuelle des corpus monolingues allemand, anglais, espagnol, français, italien et russe. Cf. http://corpora.lessicobeniculturali.net/ (22/12/2022).

[29] «Il limite che ci eravamo imposti di prendere in considerazione come prioritarie la città di Firenze e la Toscana, si sta spostando per andare a comprendere altre città italiane che hanno suscitato l'interesse dei viaggiatori stranieri nel corso dei secoli» (Farina 2020, 91).

Item	***DIV***	***LET***	***TEC***
Plaisance	5	164	-
Piacenza	-	2	-
Modene	2	156	-
Modène	13	103	4
Modena	2	9	-
Modêne	-	3	-
Ferrare	7	97	2
Ferrara	1	2	-
Reggio	5	66	-
Ravenne	13	53	2
Ravennes	-	4	-
Ravenna	1	3	-
Rimini	7	58	-
Bentivoglio	-	43	-
Bentivogli	-	1	-
Bentiviglio	-	1	-
Imola	1	40	-
Vignole	-	28	-
Vignola	-	6	-
Vignoles	-	1	-
Faenza	3	21	-
Forli	6	20	-
Forlì	3	-	-
Cento	1	16	-

« Bologne » (en fr.) est évidemment le toponyme le plus fréquent dans les trois catégories textuelles, suivi de « *Bologna* » (en it.) qui est beaucoup moins attesté[30]. La distribution différente des toponymes, en français et en italien, fournit des indices sur les stratégies et sur les approches de traduction en fonction des époques et du genre discursif, sujet qui sera repris par la suite (par. 2.2.1).

[30] Nous n'avons pas rapporté dans le tableau 3 les toponymes qui figurent dans le sous-corpus BER avec une fréquence inférieure à 15. Les autres toponymes attestés, pour la grande majorité dans les textes LET, sont: Marzabotto, Corregio / *Correges* / *Corrégio*, Cesena / *Cesene* / *Césenne* / *Césena*, Guastalla / *Guastella*, Loiano, Sassuolo / *Sassolo*, Villanova, Rubiera, Fiorenzuola, Lavino, Galliera / *Galiera*, Pianoro, Bagnacavallo, Mirandole / *Mirandola*, Saviniano, Castel-franco, Castellana, Castelvetro, Forlimpopoli, Comacchio, Rabierra / *Rubiera*, Santa-Agata / *Santagata*, Carpi, Cesenatico, Fontanellato, Fossalta / *Fossalte*, Rangone, Cicognara, Sant'Ilario, Soragna, Verucchio, Pergola, Vetulonia.

Le corpus atteste aussi l'ancienne dénomination de Bologne, « *Felsina* », qui fut la principale ville étrusque de l'Étrurie padane, et les appellation latines, « *Bononia* » et « *Bononiae* ». Après « Bologne », les villes les plus attestées sont les chefs-lieux de la région, « Parme », « Plaisance », « Modène », « Ferrare », « Ravenne », les localités touristiques les plus connues, comme « Rimini », et les villages qui se trouvent le long de l'Apennin tosco-émilien, Pianoro, Marzabotto, Loiano, etc., qui étaient des étapes pour les voyageurs du passé traversant les Apennins en carrosse ou à cheval pour rejoindre la ville de Florence à partir de Bologne et qui sont encore des destinations visées par le tourisme de montagne.

On remarque que ces toponymes sont attestés pour la plupart dans les textes LET, dans une moindre mesure dans les textes DIV et qu'ils sont presque absents des textes TEC. On peut en déduire que les textes ont été correctement classés, étant donné que les textes techniques se concentrent sur la description des œuvres d'art d'un point de vue spécialisé, et non pas sur celle des localités touristiques[31]. Nous en aurons la confirmation dans le prochain paragraphe lorsque nous analyserons les données lexico-terminologiques qui ressortent de l'examen du corpus.

3.2. Données lexico-terminologiques

Dans cette partie, nous présentons les données lexico-terminologiques qui résultent de l'analyse outillée du sous-corpus BER à l'aide du logiciel *Sketch Engine* (SE). Cette analyse, partielle et provisoire, sera approfondie lorsque ce sous-corpus sera complété avant d'être intégré dans le corpus LBC. Sa visée est, entre autres, de montrer dans quelle mesure le sous-corpus BER enrichit la couverture de la terminologie artistique du corpus LBC Français. Nous nous appuierons sur les indices donnés par deux fonctions de SE : *Keywords*, qui permet d'identifier ce qui est spécifique dans un corpus spécialisé, et *Wordlist*, ou liste de mots attestés par ordre de fréquence.

3.2.1. Keywords / Terms

La fonction *Keywords / Terms* permet d'extraire les mots-clés, ou unités lexicales simples (*single-token items*), et les termes, ou expressions polylexicales (*multi-word expressions*) qui sont uniques ou typiques du corpus BER (*focus corpus*). Ces mots et termes y apparaissent avec une plus grande fréquence par rapport à un corpus de référence (*reference corpus*), en l'occurrence le corpus français de la série TenTen (*FrenchTenTen2017*), composé de plus de 6 milliards de *tokens* issus de la Toile[32].

[31] Il pourrait y avoir des exceptions à cette règle générale : par exemple, dans les textes TEC du corpus LBC on peut trouver plus de références à Carrare lorsqu'on distingue les différents types de marbre selon leur origine.

[32] Dernière consultation le 23/07/2022.

Le tableau 4 présente la liste des mots typiques du domaine couvert par le corpus, à savoir ceux qui apparaissent dans le corpus BER plus fréquemment que dans la langue générale et qui, pour cette raison, acquièrent le statut de « termes ». On adopte ici la définition de terme telle que l'a établie L'Homme (2004, 81), la plus pertinente pour notre approche qui voit le terme inséré dans le processus discursif :

> Un terme est une unité lexicale utilisée dans un domaine spécialisé. Le lien établi entre l'unité lexicale et le domaine est central dans la démarche terminographique. [...] Bien qu'il soit appréhendé en fonction d'un repère extérieur à la langue (à savoir le domaine), le terme se matérialise dans les textes.

En nous limitant aux 30 premiers résultats, nous rapportons ci-dessous les termes qui caractérisent chaque sous-corpus, avec l'indication du *score* (*keyness score*), à savoir le résultat du rapport entre la fréquence relative d'un terme dans le corpus cible et la fréquence relative du même terme dans le corpus de référence[33]. Les termes listés sont donc les mots à forte valeur ajoutée (*high-scoring words*), c'est-à-dire les termes qui, comme l'ont expliqué Kilgariff *et al.* (2014), reflètent très bien le domaine du corpus cible et permettent de définir le thème de chaque sous-corpus.

Tab. 4. *Keywords* dans les trois sous-corpus BER.

keyword LET-BER	fréq. focus	score	keyword DIV-BER	fréq. focus	score	keyword TEC-BER	fréq. focus	score
Carrache	262	633,1	Maggiore	49	536	*Carrache*	43	**1019**
Guerchin	182	459,7	**piazza**	78	462	**villanovienne**	**26**	**667**
lieue	127	87,7	**palazzo**	64	431	*Dominiquin*	20	**503**
admirable	124	52,5	*Morandi*	32	405	Certosa	17	**430**
marbre	216	50,6	Asinelli	29	404	*Guerchin*	17	421
fresque	138	40,5	Garisenda	24	335	*Benacci*	14	465
autel	152	36,7	Petronio	24	319	**villanoviens**	**14**	**364**
tombeau	119	32,2	Archiginnasio	19	265	Etrurie	16	327
vierge	336	30,2	Poggi	18	228	**étrusque**	**26**	**280**
palais	456	25,3	*Guerchin*	16	213	*Caravage*	17	250
statue	198	23,05	*Bentivoglio*	14	185	**étrusques**	**11**	**198**
orner	145	23,02	*Bassi*	14	179	**villanovien**	**7**	**182**
tableau	808	22,75	Accursio	10	140	estudiantin	12	180
peindre	263	22,74	Nettuno	10	135	pinacothèque	9	177
mille	358	20,8	cremeria	9	127	*Reni*	7	164

[33] Nous avons éliminé les toponymes, les verbes conjugués (était/ avait, etc.) et les prénoms de personnages (Annibal, Augustin), mais nous avons gardé leurs noms de famille.

keyword LET-BER	fréq. focus	score	keyword DIV-BER	fréq. focus	score	keyword TEC-BER	fréq. focus	score
église	199	19,9	foires	11	125	ossuaire	8	134
pape	247	18,08	**museo**	15	124	**villanoviennes**	**5**	**131**
chapelle	237	17,5	podestà	9	123	terramare	5	129,7
bronze	122	16,7	erbe	9	120	<u>académisme</u>	7	129,6
peintre	233	16,3	**torri**	9	119	*Brizio*	5	129,2
cardinal	101	14,6	<u>ragù</u>	8	110	Pepoli	5	128
décorer	127	14,4	**portique**	**31**	**107**	*Gombrich*	5	125
cathédrale	134	14,3	saragozza	7	99	Prosérpine	5	105,4
ange	113	12,7	**basilique**	60	97	*Rouchès*	4	105
édifice	117	12,5	cavour	9	96	Pigorini	4	103
gloire	102	11,5	<u>tortellini</u>	7	95	<u>biconique</u>	4	102,8
galerie	144	11,1	podestat	8	94	*professore*	4	102
bâtir	113	10,89	*Caravage*	11	87	*Lavergnée*	4	101
<u>beauté</u>	211	10,81	<u>tagliatelle</u>	9	85	*Brejon*	4	100
<u>considérable</u>	101	10,81	Farini	6	83	<u>esthète</u>	6	99

Le premier terme listé dans les sous-corpus LET et TEC, avec un *score* très élevé, est « Carrache » (*Carracci*), le nom de famille des deux frères Annibal et Augustin et de leur cousin Ludovic, les principaux représentants de l'école bolonaise de peinture. On en déduit que leurs œuvres, ainsi que celle de Guerchin, nom qui est aussi très attesté, font l'objet de commentaires et de descriptions dans les textes de ces deux sous-corpus en particulier.

Le sous-corpus LET présente, comme mots-clés, des noms propres d'artiste (*Carrache, Guerchin*) et un grand nombre de termes des trois domaines pris en compte (en gras dans la première colonne du tableau 4). Le thème dominant de ce sous-corpus DIV est donc le patrimoine artistique tout court. On remarque aussi la présence des adjectifs évaluatifs, « admirable », « considérable » (soulignés), qui confirment que les textes témoignent du regard subjectif des visiteurs étrangers face au patrimoine artistique bolonais. Ces données montrent la pertinence des choix méthodologiques effectués par les concepteurs du projet LBC d'intégrer un grand nombre de textes littéraires pour attester le lexique employé pour le décrire.

Les textes DIV présentent des caractéristiques bien différentes. Ici très peu de termes du domaine de l'art (en gras), alors que les mots-clés avec le *score* le plus élevé sont les noms propres des monuments et des lieux symboles de la ville de Bologne : Maggiore, le nom de la place principale, Garisenda et Asinelli, les deux tours les plus connues, les noms des palais (Poggi, d'Accursio), des musées (Archiginnasio), des églises (San Petronio) et des *piazze* (Nettuno). Ce qui ressort ici, c'est la présence de noms appartenant au champ lexical de la gastronomie (soulignés) qui désignent les produits typiques de cette région (*ragù, tortellini, tagliatelle*),

ce qui confirme que, la nourriture faisant partie du patrimoine culturel partagé, ce corpus a été construit en tenant compte d'une définition large de patrimoine. Les mots-clés indiquent ainsi que ces textes déploient des informations sur les curiosités touristiques et gastronomiques de la ville de Bologne. La présence abondante de termes en italien pour désigner les places, les palais, les musées et même les tours confirme ce que nous avons observé concernant les toponymes : dans les textes de vulgarisation on a plus recours à l'italien que dans d'autres typologies textuelles, ce qui rejoint l'observation de Farina et Billero (2020, 172) : « The abundant use of Italianisms is a dominant feature of the tourist guides analyzed ».

Les textes TEC présentent des mots-clés qui se distinguent nettement des deux autres sous-corpus. Le score le plus élevé est celui qui résulte de la somme des différentes formes fléchies de l'adjectif « villanovien ». Le thème dominant est ici la protohistoire bolonaise, ce qui est corroboré par la présence des mots-clés « Benacci »[34], « Étrurie », « étrusque », « ossuaire », « terramare »[35]. On remarque aussi la présence d'un terme du domaine de l'archéologie (« biconique ») et de quelques termes qui relèvent de la critique d'art (« académisme », « esthète »). Parmi les noms propres, outre les noms d'artistes de l'école bolonaise du XVIe siècle (Carrache, Dominiquin, Guerchin, Reni, Brizio) et du milanais Caravage, on relève la présence de noms d'historiens de l'art du XXe siècle (Gombrich, Rouchès, Brejon de Lavergnée) et d'un archéologue (Pigorini). En raison de sa très petite taille, le sous-corpus TEC permet ainsi de recueillir, par bribes, des attestations à la fois de la terminologie scientifique et technique des domaines concernés, mais aussi de la terminologie de la critique d'art qui se caractérise par un plus grand degré de liberté dans l'expression dans le but de surprendre le lecteur. C'est principalement sur l'ajout de textes techniques d'abord, et de textes de vulgarisation, que portera le travail d'équilibrage du corpus.

L'extraction des expressions polylexicales (*terms* ou *multi-word expressions*), qui sont uniques ou typiques du corpus BER (*focus corpus*), confirme les observations que nous venons de formuler. Nous les énumérons ci-après, sans en préciser la fréquence et les *scores* par souci de brièveté :

- *Terms* LET : *Louis Carrache ;*
- *Terms* DIV : Piazza Maggiore, San Luca, San Petronio, Tour Asinelli, Santo Stefano, Giorgio Morandi, Tour Garisenda, Basilique San Petronio, Mercato delle Erbe, *Laura Bassi*, Palazzo Poggi;
- *Terms* TEC: école bolonaise, archive historique, *Annibal Carrache*, Première tombe, Tombe Benacci, *Guido Reni*, Palais Pepoli, Sainte Cécile, Art photographique, Tableau bolonais, Tombes villanoviennes.

Nous concluons que la méthodologie offerte par l'analyse statistique des corpus, notamment par la fonction *Keywords/Terms* de SE, s'avère très performante

[34] Nom d'un tombeau de la Bologne étrusque.

[35] Nom d'un peuple qui habitait, avec les Villanoviens, la région de Bologne durant la préhistoire (cf. Cotteau 1889).

pour la compréhension du contenu et pour l'analyse préliminaire des termes les plus significatifs du sous-corpus. La *Wordlist*, qui fera l'objet du paragraphe suivant, permet de compléter cet aperçu avec l'analyse des fréquences de tous les mots attestés dans le corpus.

3.2.2. Wordlist

Nous nous sommes concentrée sur deux catégories grammaticales, les verbes et les noms. Les tableaux 5 et 6 rapportent la liste, respectivement, des formes lemmatisées des verbes et des noms dont la fréquence est égale ou supérieure à 20 dans un des trois sous-corpus BER. Nous avons trié manuellement les résultats de l'extraction automatique faite par SE, en supprimant de la liste des noms les toponymes, déjà vus plus haut, les noms propres des artistes (Carrache, Guerchin, etc.), les noms qui relèvent des professions (maître, peintre, artiste, etc.), des matériaux (marbre, pierre), des parties du corps (main, tête, pied), et les termes abstraits (beauté, goût, etc.). Nous avons également éliminé tous les noms et les verbes qui ne relèvent pas strictement du discours sur l'art (« école », « université », « étudiant », « visite », « visiter », « marcher », etc.).

Tab. 5. *Wordlist* des verbes dans le sous-corpus BER.

Item	*fréq. LET*	*fréq. DIV*	*fréq. TEC*
représenter	404	32	58
peindre	323	14	21
élever	249	16	10
conserver	196	30	30
composer	191	11	7
orner	176	7	2
bâtir	**163**	10	2
décorer	151	4	4
construire	116	34	4
dessiner	95	1	2
exécuter	91	1	5
<u>**étayer**</u>	73	17	20
graver	48	2	2
dorer	42	1	3
garnir	39	3	2
ériger	29	8	1
sculpter	28	6	1
paver	26	2	1
restaurer	24	14	5
fortifier	23	-	-
embellir	21	-	2

Tab. 6. *Wordlist* des noms dans le sous-corpus BER.

item	*fréq. LET*	*fréq. DIV*	*fréq. TEC*
tableau	808	6	78
piazza	**13**	**78**	-
palais	**456**	**77**	**15**
église	610	100	15
ouvrage	334	6	3
vierge	330	11	-
maison	317	44	17
figure	292	6	20
peinture	259	31	74
place	238	69	12
chapelle	237	18	2
statue	198	20	1
couleur	183	17	14
tour	174	122	12
porte	170	31	3
cour	**169**	**43**	**26**
autel	152	5	-
galerie	**144**	**14**	**14**
fresque	138	21	4
cathédrale	134	11	1
portique	112	31	1
composition	119	4	14
tombeau	119	15	-
édifice	117	31	14
colonne	94	11	2
fontaine	84	20	2
musée	51	**76**	**45**
loge	47	-	**13**
coupole	72	3	-
pont	72	2	3
dessin	72	15	27
arcade	84	45	5
coloris	71	1	-
palazzo	24	64	3
rue	159	60	4
quartier	15	51	-
paysage	39	8	56
tombe	7	9	39
restauration	4	17	27
technique	1	2	22

Dans le tableau 6, on observe que certains concepts sont désignés en langue italienne surtout dans les textes du vulgarisation (*piazza* : 78 occurrences dans DIV *vs* 13 occurrences dans LET ; *palazzo* : 64 occ. dans DIV *vs* 24 dans LET et 3 dans TEC), alors que leur équivalent en français figure plus fréquemment dans les textes littéraires (« place » : 238 occurrences dans LET *vs* 69 dans DIV et 12 dans TEC; « palais » : 456 occ. dans LET *vs* 77 occ. dans DIV et 15 dans TEC). Cette donnée, qui est observable aussi dans la liste des toponymes (cf. Tab. 3 dans paragraphe 2.1.3), signale que la dénomination d'un concept peut varier (dans ce cas en français ou en italien) selon la typologie textuelle et selon le pacte de lecture avec le destinataire. Un corpus comparable spécialisé, qui rend possible la recherche par métadonnées des catégories et sous-catégories textuelles (Billero 2020, 30), donne ainsi des indices sur les stratégies de traduction du lexique qu'il faudra déployer en tenant compte de ces variables.

Pour ce qui concerne la couverture de la terminologie artistique, mise à part la présence de certains hyperonymes pouvant se rapporter aux trois domaines confondus (« représenter », « composer », « orner », « embellir » parmi les verbes et « ouvrage », « figure », « composition », « restauration », « technique » parmi les noms) ou à des sujets représentés dans tous les arts (« vierge »), ainsi que du phénomène de la polysémie (« composer »), on remarque que chacun des domaines pris en considération(architecture, peinture, sculpture) est bien représenté par du « lexique charnière » (Cabré 1991), à savoir un lexique puisé dans la langue commune et employé aussi dans un domaine spécialisé, et rarement par des termes hautement spécialisés, cela principalement à cause du fait que, comme on l'a vu, les textes techniques sont à l'état actuel très peu représentés :

- Pour la peinture : « peindre », « dessiner » ; « tableau », « couleur », « fresque », « composition », « dessin », « paysage », « coloris » ;
- Pour l'architecture : « élever », « bâtir », construire », « étayer », « ériger », « fortifier » ; « *piazza* », « palais », « église », « maison », « place », « chapelle », « tour », « porte », « cour », « autel », « galerie », « cathédrale », « édifice », « portique », « musée », « coupole », « pont », « arcade », « *palazzo* », « rue » ;
- Pour la sculpture : « orner », « graver », « sculpter » ; « statue », « tombeau », « colonne », « fontaine », « tombe ».

On relève dans le tableau 5 la présence d'un terme technique du domaine de l'architecture, le verbe « étayer », qui est en fait attesté dans le *Grand Dictionnaire Terminologique* (GDT) comme relevant de l'industrie de la construction[36]. Il est intéressant de comparer les données relatives à sa fréquence : malgré le fait que le sous-corpus TEC a une taille plus petite que le sous-corpus DIV, ce verbe y est plus fréquent (20 occurrences dans TEC *vs* 17 dans DIV). La même remarque peut se rapporter à d'autres verbes comme « peindre », « exécuter » », « repré-

[36] Cette donnée ouvre des questionnements sur la représentation actuelle du domaine des beaux-arts dans les banques de terminologie.

senter » qui sont aussi plus fréquents dans les quelques textes TEC que dans les textes DIV (cf. Tab. 5). Cela suggère que dans DIV le patrimoine artistique est simplement mentionné et non pas décrit de manière pointue.

On en déduit que la quantité de lexique charnière relatif au domaine des arts est plus importante dans les textes LET et TEC, alors que dans les textes DIV, c'est plutôt le lexique de la langue générale qui est le plus représenté. Cet acquis, bien qu'il repose sur une quantité de données qui n'est pas suffisante ici, témoigne de la relation étroite entre le contenu des trois sous-corpus pris singulièrement et le degré de spécialisation de la langue attestée, qui corrobore encore une fois la pertinence des choix méthodologiques adoptés par les concepteurs du projet LBC : les textes DIV représentent la langue générale ou non spécialisée, le lexique charnière ou sub-technique est incarné par le sous-corpus LET et le lexique spécialisé sera attesté par l'élargissement du corpus TEC.

Les tableaux 5 et 6 affichent une autre donnée qui montre distinctement la spécificité du corpus BER et sa complémentarité avec le corpus LBC Français : l'échantillon de noms et verbes relatifs au domaine de l'architecture est le plus riche. Cela dépend sans doute du fait que, d'une part, le patrimoine artistique de la ville de Bologne compte une grande partie de biens immobiliers (palais, musées, édifices historiques) appartenant à l'Université, et que, d'autre part, un patrimoine architectural très important pour cette ville est représenté par les « arcades » (ou « portiques »), nommés par l'UNESCO en 2021, qui sont, avec les nombreuses anciennes « tours », un symbole de la ville et de la région. Elles font en effet l'objet d'un très grand nombre d'attestations dans toutes les catégories textuelles (Tab. 5). On approfondira la portée de ces éléments dans la prochaine section, consacrée à l'application de la méthode d'analyse choisie.

4. Stratégies de traduction basée sur corpus (CTS) : résultats significatifs du corpus BER

Dans cette dernière section, nous essayerons d'illustrer dans quelle mesure l'approche adoptée dans le cadre du projet LBC (*corpus-driven*) permet de « faire émerger de manière inductive des savoirs linguistiques » (Williams 2005, 13), concernant notamment le phénomène de la synonymie (diastratique et diatopique) en terminologie. On montrera que ces savoirs, induits à partir d'un corpus comparable, fournissent des pistes pour déployer de nouvelles stratégies pour la traduction du lexique artistique qui ne sont pas prises en compte par les ressources plurilingues disponibles à l'heure actuelle. On se concentrera sur un cas d'étude : la traduction française des *portici* bolonais.

4.1. Cas d'étude : la traduction française des *portici* bolonais

L'analyse rapprochée des données lexico-terminologiques du sous-corpus BER révèle une distribution différente, dans les trois typologies textuelles, d'une série de synonymes terminologiques employés en langue française pour désigner les célèbres *portici* (en italien) de la ville de Bologne. Ce sont : « portiques »,

« arcades » et « galerie »[37]. Nous avons établi cette série synonymique en nous fondant, dans un premier temps, sur la *wordlist* et sur l'analyse des concordances et, dans un second temps, sur la lecture attentive des extraits textuels. Ces deux passages ont confirmé que ces mots sont bien employés dans les textes français comme synonymes intralinguistiques du mot italien *portici*. Nous en rapportons ci-dessous un échantillon représentatif :

> Ce que Bologne offre de vraiment singulier en édifices, c'est cette **galerie**[38] composée de sept cents arcades, fermée au nord, ouverte au midi et qui, dans la longueur d'une lieue, conduit de la porte de la ville à celle d'une église [...] L'église à laquelle aboutit ce **portique**, est tellement couverte d'*ex voto*, qu'on en remarque à peine toutes les proportions. (LET, La Porte 1779)

> Pour rendre ce pèlerinage facile & commode, même agréable en tout temps, il a été construit une **Galerie**, ou **Portiques ouverts**, qui partent des murs de la ville, & conduisent sans interruption jusqu'au sol & près du portail de la Madonna : cette distance distribue au delà de sept cents cinquante arcades d'une construction uniforme, agréable & solide. (LET, La Roche 1783)

> Près de la chartreuse s'ouvre l'entrée de cet immense et singulier **portique** qui fut commencé en 1674, fini en 1739, au moyen d'une contribution volontaire des habitans de Bologne, et dont les six cent trente-cinq arches courent sur une ligne de près de 5 kilomètres, d'abord en plaine, puis sur la pente du Monte della Guardia. (LET, Rémusat 1861)

> Comme il y a dans cette chartreuse un grand nombre de cloîtres au rez-de-chaussée, dont chacun présente un parterre et quatre faces de galerie, le cimetière pourra suffire très longtemps aux besoins de la ville, à laquelle on se propose de le joindre prochainement, par une suite d'**arcades** de même espèce que celles qui bordent, de chaque côté, les plus belles rues de Bologne, et qui y sont d'une grande ressource pour les gens de pied contre le soleil et contre la pluie. (LET, Plessis 1903)

> L'église *Madonna San Luca*, sise au pied des monts, tout près de Bologne, est un lieu de pèlerinage pour tous les fidèles de l'Italie. On s'y rend en passant sous une **arcade** de 640 arches magnifiques. (LET, Couture 1926)

> La *Madonna di S. Luca*, église située à une lieue de Bologne sur une montagne ; on y va par une belle **galerie en portiques**, formée de 700 arcades, et qu'on appelle il *Porticato di S. Luca*, qui rend le pélerinage très-commode : c'est un des plus grands monumens de la dévotion des Italiens à la Sainte Vierge : mais il ne doit son existence qu'à la générosité des habitans de Bologne. (La Lande 1769 chap. VI)

[37] Il y a aussi une occurrence de « arches » pour désigner les « arcades » dans Rébéna (2017).

[38] Le gras est le nôtre.

> À une lieue de Bologne, et sur la montagne de Guardia, est une église dédiée à la Sainte-Vierge, et où l'on arrive par un **portique** de 690 arceaux qui commence à la porte de la ville et va jusqu'au sommet de la montagne. (LET, Starke 1833-1834)

> À l'ouest est le *monte della Guardia* précédé d'un long **portique** de 650 arcades. Au centre deux hautes et antiques tours, penchées d'une manière effrayante. (LET, Marmottan 1918)

> Parmi les coupoles et les places, l'arc de Meloncello mène à un sanctuaire par le passage sous un grand **portique**. Une trentaine d'arcades sur cent quarante mètres de façade se succèdent, soutenues par des colonnes doriques. (DIV, La Presse 1968)

> Tout en cheminant lentement et en m'arrêtant devant les curiosités de la ville, me voici bientôt en face d'un immense **portique** en maçonnerie composé de centaines d'arcades (630), la plupart ornées de peintures et de pieuses inscriptions. Cette **galerie** qui s'étend sur une ligne de près de quatre mille, monte sur la colline jusqu'à l'église Saint-Luc. (LET, QB, Rocheleau 1922)

> Les 666 **arcades de voûte**, d'une longueur de près de quatre kilomètres (3796 m), relient efficacement San Luca, comme on l'appelle communément, au centre-ville. Ces **portiques** abritent la procession traditionnelle qui porte chaque année depuis 1433 une icône byzantine de la Vierge à l'Enfant attribuée à Luc l'évangéliste jusqu'à la cathédrale de Bologne pendant la fête de l'Ascension. (DIV, Buchot 2020)

L'observation de l'emploi en discours de données langagières authentiques à l'intérieur du corpus prouve qu'il est donc possible de désigner en français les « *portici* » bolonais en faisant appel à différents moyens lexicaux et discursifs (signalés en gras). Si on transpose ces savoirs à la sphère de la traduction, on retiendra que plusieurs équivalents traductionnels de « *portici* » sont envisageables : « galerie », « portique », « arcade », au singulier ou au pluriel[39], ainsi que « portique ouvert », « galerie ouverte », « galerie en portiques », « arcades de voûte ».

En contexte multilingue, l'objectif de cette méthode d'analyse serait idéalement de repérer des équivalents traductionnels de mots dans des corpus comparables, permettant de pallier le manque de corpus parallèles (Zweigenbaum, Habert 2006). L'absence à l'état actuel d'un corpus monolingue BER italien ne nous permet pas de comparer ces données avec un ensemble de textes qui, dans la langue italienne, sans être en rapport de traduction mutuelle, traitent des mêmes sujets et domaines couverts par le corpus BER. Cependant, nous pouvons accéder à des textes originaux produits en français qui suggèrent des équivalents traductionnels potentiels dont les ressources lexicographiques d'aide à la rédaction plurilingue et à la traduction ne font pas état. Nous montrons, à titre d'exemple, l'entrée *portico* du dictionnaire bilingue de référence français-italien

[39] Le mot « arcades » au pluriel se réfère dans plusieurs cas aux arches dont les portiques sont composés. Dans ces cas, nous ne l'avons pas signalé en gras.

Boch (2020) de la maison Zanichelli, où trois équivalents traductionnels de ce mot sont proposés : 1. « portique », 2. « arcades » et 3. « porche » (Fig. 2).



s. m. [pl. *-ci*]

1 **portique**: *i portici di Atene*, les portiques d'Athènes

2 (*di strade, piazze e sim.*) **arcades** (f. pl.): *passeggiare sotto i portici*, se promener sous les arcades

3 (*di chiese, edifici e sim.*) **porche**

4 **porche**: *il portico di una fattoria*, le porche d'une ferme.

Fig. 2 Entrée « portico » dans le dictionnaire Boch (2020).

Les informations métalinguistiques présentées entre parenthèses avant les équivalents 2 et 3 signalent, respectivement, que « arcades » au pluriel ne serait utilisé qu'en référence, ou en collocation, avec des rues, places et entités similaires (*di strade, piazze e sim.*) et que « porche » serait employé pour les *portici* des églises et édifices de toutes sortes. Aucune information n'est donnée par ailleurs pour guider l'utilisateur dans la compréhension de l'emploi de l'équivalent « portique ». Au contraire, la présence d'un seul exemple relatif aux portiques d'Athènes pourrait le dérouter, le conduisant à considérer que ce mot ne peut se rapporter qu'à l'architecture grecque.

Or, ces informations ne coïncident pas, voire sont en contradiction avec les données relevées dans le sous-corpus BER, étant donné que « arcades » est bien attesté pour désigner les *portici* longeant les rues de Bologne (Plessis 1903), et que c'est « portique », et non pas « porche » (très peu attesté dans le sous-corpus BER)[40], qui est principalement employé en collocation avec des églises, notamment l'église de la Madonna di San Luca (La Porte 1779, Starke 1883, etc.). En outre, « galerie », un autre synonyme largement attesté dans le corpus BER, n'est pas du tout pris en compte par ce dictionnaire.

Une analyse plus approfondie des données relatives à la fréquence statistique de ces synonymes intralinguistiques nous permet d'estimer que les lacunes des ressources existantes portent non seulement sur la couverture des équivalents traductionnels possibles, comme on vient de le voir, mais aussi sur la prise en compte de leur emploi en discours. En fait, l'examen des données statistiques relatives à la fréquence et à la distribution de ces synonymes dans les trois caté-

[40] 3 occurrences de « porche » et dans BER-DIV et dans BER-LET.

gories textuelles du corpus BER nous amène à formuler un certain nombre de remarques sur la variabilité de l'emploi de ces mots / termes en discours.

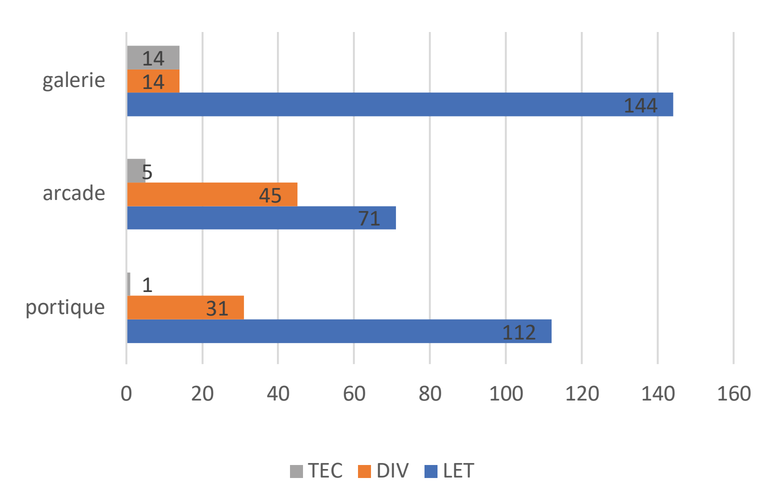


Fig. 3 Occurrences de « portique », « arcade » et « galerie » dans les sous-corpus BER.

Compte tenu du fait que les textes DIV ne représentent que 12% du sous-corpus contre 80% des textes LET (Fig. 3) :

- la fréquence de « arcade » (45) y est proportionnellement beaucoup plus élevée par rapport au texte LET (71), à savoir ce terme est beaucoup plus employé dans les textes DIV que dans les textes LET ;
- « portique » (94) est plus fréquent que « arcade » (71) dans les seuls textes LET ;
- « arcade » (45), au singulier et au pluriel, est plus fréquent que « portique » (31) dans les textes DIV;
- « galerie », souvent employé comme synonyme hyperonymique de ces deux termes, est bien plus fréquent dans les textes littéraires (144) que dans les textes DIV (14) et TEC (14).

Ces données semblent confirmer « l'existence de nombreux synonymes dans la terminologie artistique, qui ne témoignent pas de différences de nature conceptuelle mais diastratique » (Cetro 2022, 140). On suppose donc que dans la langue générale, représentée par le sous-corpus DIV, c'est le mot « arcades » qui traduit le mot italien *portici*, bien que la traduction « portique » soit également possible mais avec une fréquence inférieure. Dans la langue sub-technique, représentée par le sous-corpus LET par la voix d'un public cultivé mais non spécialiste, c'est en revanche le mot « portique » qui traduirait plus fréquemment le mot italien *portici*. Quant au discours spécialisé de l'architecture, nous n'avons pas suffisamment d'attestations de ces termes dans le sous-corpus TEC pour pouvoir avancer des hypothèses sur l'emploi de ce terme. On relève

toutefois que la comparaison de la fréquence de cette série synonymique dans le sous-corpus BER avec le corpus LBC Français fournit d'autres indices intéressants qui jettent une lumière sur son emploi dans le discours spécialisé, voire dans la lexicographie de spécialité[41].

En tenant dûment compte des différentes proportions du corpus LBC Français et du sous-corpus BER, étant donné que la taille de ce dernier correspond à 1/6e de la taille totale du corpus LBC Français, le premier indice intéressant est le constat qu'il n'y a pas d'écart considérable entre le nombre d'occurrences de « arcade » (151) et de « portique » (123) dans le sous-corpus LET LBC Français, alors que cet écart est bien plus évident dans le sous-corpus LET BER Français, où la fréquence de « portique » (112) est plus élevée que celle de « arcade » (71) et proportionnellement beaucoup plus importante que dans le corpus LBC (cf. Figg. 4 et 5).

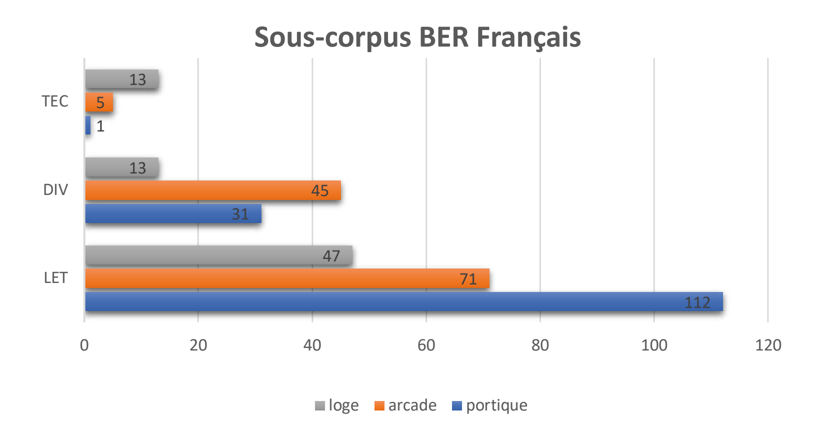


Fig. 4 Nombre d'occurrences de « arcade », « portique » et « loge » dans le corpus BER.

On remarque aussi que dans les textes techniques du corpus LBC, « portique » est beaucoup plus fréquent (257) que dans les textes LET (123), ce qui suggère que l'utilisation de « portiques » pourrait avoir droit de cité dans des textes écrits par des professionnels du domaine. Tout cela nous amène à conclure que l'hypothèse avancée à partir de la seule consultation du corpus BER est validée par la consultation du corpus LBC Français, qui porte principalement sur la Toscane, à savoir que « portique » est bien un terme du domaine de l'architecture. Le fait qu'il figure, à l'intérieur du corpus LBC, aussi bien dans un dictionnaire spécialisé, le *Dictionnaire raisonné de l'architecture française* d'Eugène Viollet-le-Duc (1854-1868), que dans un ouvrage de l'historien de l'art Léon

41 Nous avons exclu de cette comparaison le terme « galerie », bien que fréquent dans le sous-corpus, à cause de sa polysémie qui infirme certaines considérations fondées ici sur des critères purement statistiques.

Palustre (1892), confirme qu'on se trouve bien face au phénomène de la synonymie diastratique en terminologie.

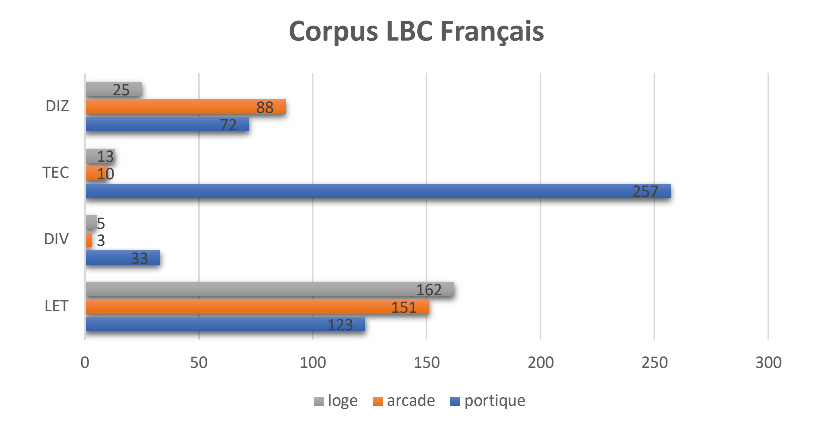


Fig. 5 Nombre d'occurrences de « arcade », « portique » et « loge » dans le corpus LBC Français.

« Portique » est un terme, très employé pour désigner les « *portici* » bolonais, comme on l'a déduit de sa fréquence proportionnellement très élevée dans les textes littéraires et de vulgarisation du corpus BER. Ces données nous amènent à émettre une autre hypothèse : « portique », un calque de la langue italienne, peut être aussi considéré comme un mot désignant une *realia*, une réalité typique de l'architecture bolonaise. Autrement dit, un locuteur français non spécialiste emploierait « arcades » pour désigner toute « construction formée d'un arc de voûte soutenu par des piliers ou des colonnes » (TLFi), y compris les « arcades » de la rue de Rivoli ou de la Place des Vosges à Paris. En revanche, un autre type de locuteur, à qui l'architecture bolonaise serait familière sans être un spécialiste, comme nos voyageurs francophones représentés par le sous-corpus BER-LET, emploierait par osmose le calque « portique » qui désigne précisément l'architecture culturo-spécifique bolonaise et non pas 'tout type' d'« arcades » (cf. Zotti 2003). La myriade d'articles disponibles sur la Toile qui annoncent la nomination des « portiques » de Bologne au patrimoine culturel de l'UNESCO semblent confirmer le constat qui émerge de l'analyse du corpus.

Pour finir, la comparaison des données de BER avec le corpus LBC Français a fait apparaître une autre donnée significative : « loge » et l'italianisme « *loggia* », tous les deux assez fréquents dans le Corpus LBC Français (« loge » : 162 dans LET ; *loggia* : 110 au total), sont en revanche presque absents du sous-corpus BER (« loge » : 47 dans LET et 13 dans TEC et DIV ; *loggia* : 3 au total)[42].

[42] Dans l'édition italienne de 1568 des *Vite* de Vasari, il y a de nombreuses occurrences du mot *loggia* en italien. Les traductions françaises que nous avons analysées en vue de la consti-

La fréquence et la distribution différentes de ces synonymes partiels refléteraient aussi à notre avis l'influence des facteurs géographiques et historiques sur la dénomination des concepts, étant donné que diverses traditions architecturales dans les régions d'Italie correspondent à une diffusion variée des termes les dénommant. On voit donc clairement que les corpus sont affectés par les différences linguistiques entre les cultures (cf. Laviosa 2002), et que l'histoire de la terminologie des arts et métiers permet de pénétrer dans l'histoire des sociétés, en particulier européennes (cf. Zanola 2018). Cette affirmation est tout à fait pertinente même lorsqu'on se réfère à un pays comme l'Italie, longtemps fragmenté politiquement et culturellement.

La consultation d'une ressource de dernière génération, *BabelNet*[43], présentée comme « le plus grand dictionnaire encyclopédique multilingue continuellement mis à jour » (Navigli 2018) et qui est, de fait, un réseau sémantique multilingue et une ontologie lexicalisée, ne donne pas de résultats plus exhaustifs que le dictionnaire bilingue Boch-Zanichelli, tout au moins pour ce qui concerne la traduction du lexique artistique (Fig. 6).

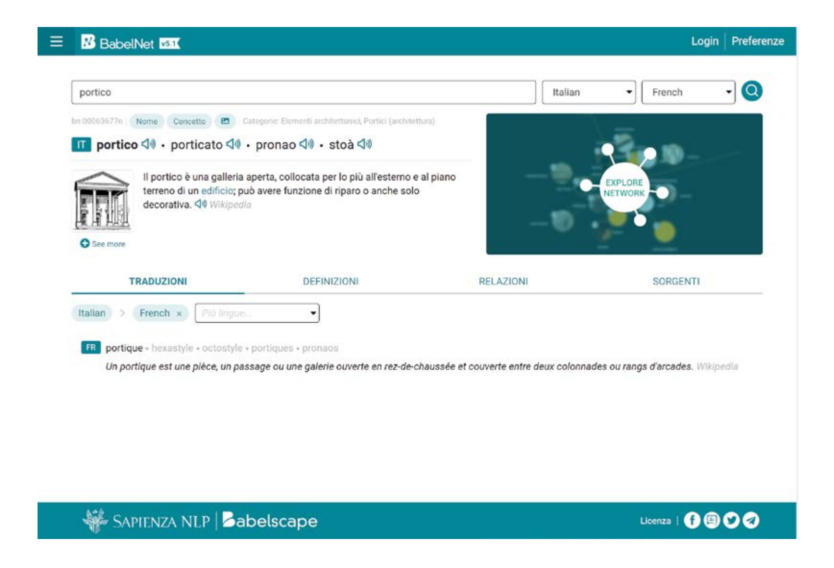


Fig. 6 L'entrée italien français « portico » dans BabelNet.

On considère de la sorte que l'intégration d'un sous-corpus sur le patrimoine de l'Émilie-Romagne à l'intérieur du corpus LBC, qui se focalise principalement sur la Toscane, pourrait révéler aussi des informations inédites sur les synonymes

tution d'un corpus parallèle des traductions de cet ouvrage attestent ces trois équivalents traductionnels : *portiques*, *galerie* et *loggia*.

43 Cette ressource, disponible sur https://babelnet.org/ (22/12/2022), a été développée par le groupe *Natural Language Processing* de l'Université Sapienza de Rome.

terminologiques de nature diatopique qui auront également des retombées sur l'attribution de l'équivalent le plus pertinent dans la pratique traductive. L'approche proposée par la traduction basée sur les corpus (*Applied Corpus Translation Studies*) s'avérera donc précieuse pour traduire plus précisément des termes du patrimoine de la ville de Bologne à l'intérieur des fiches du futur dictionnaire plurilingue du patrimoine culturel LBC (cf. Farina, Billero 2020).

5. Conclusions

La ville de Florence a été la destination privilégiée de nombreux voyageurs aristocrates européens du Grand Tour, elle a continué de l'être au courant des XIXe et XXe siècles et l'est encore aujourd'hui. En revanche, la ville de Bologne n'a généralement été qu'une étape de passage pour les voyageurs qui se dirigeaient vers Florence et Rome, après avoir traversé les Apennins Tosco-Émiliens. En conséquence, Bologne a toujours bénéficié d'un type de tourisme de nature différente, davantage « culturel » qu'artistique, à cause de la présence de son ancienne Université, et, plus récemment, grâce à sa nomination comme ville créative UNESCO en 2006 et à l'insertion de ses *portici* dans la liste du Patrimoine Mondial UNESCO en 2021. Il va de soi que les récits de voyage, les guides touristiques et les œuvres critiques du passé et du présent n'ont pas été publiés dans la même proportion que pour la ville berceau de l'art italien et que les auteurs ne se sont pas arrêtés sur les mêmes thématiques. Cette diversité constitue pourtant un atout pour le futur élargissement du corpus LBC.

La constitution du sous-corpus Bologne et Émilie-Romagne va, de fait, dans la direction du premier objectif visé par l'équipe de recherche LBC, à savoir la constitution d'un corpus comparable multilingue centré sur la description du patrimoine artistique italien, étape nécessaire et préalable à la création d'un dictionnaire plurilingue du patrimoine culturel. L'exploitation d'un corpus comparable de ce type, qui prend en compte un large éventail d'énonciateurs et de situations dans lesquelles le discours plus ou moins spécialisé sur l'art italien est produit, présente un grand intérêt aussi bien dans des travaux de traduction que dans des recherches en terminologie et en sociolinguistique, comme nous l'avons vu en nous arrêtant, avec notre cas d'étude, sur le phénomène de la variation diastratique et diatopique en terminologie. Ce simple cas d'étude, qui sera approfondi par la suite, a dévoilé les deux « âmes » de la terminologie artistique, l'âme humaniste et l'âme technique, tout en jetant une lumière aussi bien sur l'histoire des langues et des cultures que sur les dynamiques de la communication spécialisée selon une perspective à la fois nationale et internationale.

Le corpus comparable multilingue LBC a la vocation d'être une ressource flexible, créée *ad hoc* pour une étude ciblée du lexique artistique. Sa taille, déterminée par la finalité du corpus, est modulable et en continuelle expansion. Le logiciel d'exploration adopté par le projet LBC permet d'exploiter tout son potentiel, car la recherche par métadonnées permet d'obtenir des informations ciblées sur l'emploi des mots en fonction du contexte et du genre textuel, d'extraire de la terminologie et, non moins important, d'améliorer des compétences

en traduction et en langue étrangère. Nous avons en effet bien vu que, puisque le degré de technicité et de spécialisation des corpus comparables multilingues peut varier de la langue générale aux langues spécialisées, tant les chercheurs que les enseignants et les étudiants peuvent les employer comme des ressources complémentaires aux dictionnaires et à d'autres matériels linguistiques (cf. Zanettin 1998).

Dans la prochaine étape de notre recherche, centrée sur la création de bases parallèles de traduction de l'italien vers le français des *Vies* de Georges Vasari, il sera possible de mettre en perspective les avantages offerts par les corpus comparables multilingues avec les points de force des corpus parallèles en diachronie. C'est à travers l'analyse des traductions publiées au cours des siècles que l'évolution de la terminologie artistique en français sera mise en lumière. Elle pourra ainsi répondre aux nécessités en matière de traduction spécialisée multilingue et appréhender des savoirs d'un très vaste patrimoine historique et culturel, tout cela en continuité avec la vocation et les objectifs visés par les deux projets dans lesquels l'équipe LBC-CeSLiC de l'Université de Bologne est impliquée.

Références bibliographiques

Baker, Mona. 1993. "Corpus Linguistics and Translation Studies. Implications and Applications." In *Text and Technology: In honour of John Sinclair*, a cura di Gill Francis, Bill Louw, Alan Partington ed Elena Tognini-Bonelli, 233-50. Amsterdam-Philadelphie: Benjamins.

Bernardini, Silvia, e Adriano Ferraresi. 2013. "Old Needs, New Solutions: Comparable Corpora for Language Professionals." In *Building and Using Comparable Corpora*, a cura di Serge Sharoff, Reinhard Rapp, Pierre Zweigenbaum e Pascale Fung, 303-19. Berlin: Springer.

Biber, Douglas. 2012. "Corpus-based and Corpus-driven Analyses of Language Variation and Use." In *The Oxford Handbook of Linguistic Analysis*, a cura di Berndt Heine e Heiko Narrog, 159-92. Oxford: Oxford University Press.

Billero, Riccardo. 2020. "Implementazione di software per la gestione dei corpora LBC." In *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, a cura di Riccardo Billero, Annick Farina e María Carlota Nicolás Martínez, 19-31. Firenze: Firenze University Press.

Billero, Riccardo, e María Carlota Nicolás Martínez. 2018. "Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC." *CHIMERA: Romance Corpora and Linguistic Studies* 4(2): 203-16.

Boch, Raoul. 2020. *Le Robert & Zanichelli il Boch : Dictionnaire français-italien et italien-français*. Bologna: Zanichelli.

Cabré, Maria Teresa. 1991. "Terminologie ou terminologies ? spécialité linguistique ou domaine interdisciplinaire ?" *Meta. Journal des traducteurs / Translators' Journal* XXXVI: 55-63.

Cetro, Rosa. 2022. *La démarche terminologique d'André Félibien. La systématisation du lexique artistique*. Torino: L'Harmattan Italia.

Cetro, Rosa, e Valeria Zotti. 2020. "Les corpus et la base terminologique LBC: des ressources pour la traduction du patrimoine artistique." In *Lexique(s) et genre(s) textuel(s) : approches sur corpus. Actes de la conférence 11es journées du réseau*

Lexicologie, Terminologie, Traduction, a cura di Mathieu Mangeot e Agnès Tutin, 81-98. Paris: Éditions des Archives Contemporaines.

Colson, Jean-Pierre. 2008. "Cross-linguistic phraseological studies: An overview." *Phraseology. An interdisciplinary perspective*, a cura di Sylviane Granger e Fanny Meunier, 191-206. Amsterdam-Philadelphia: Benjamins.

Colson, Jean-Pierre. 2015. *The contribution of corpus-based phraseology to translation studies: from experiments to theory*. Europhras (Malaga, 29/06/2015 - 01/07/2015). http://hdl.handle.net/2078.1/165297 (22/12/2022).

Cusatelli, Giorgio. 1986. *Viaggi e viaggiatori del Settecento in Emilia e in Romagna*. Bologna: Il Mulino.

Farina, Annick, e Riccardo Billero. 2018. "Comparaison de corpus de langue "naturelle" et de langue "de traduction": les bases de données textuelles LBC, un outil essentiel pour la création de fiches lexicographiques bilingues." In *JADT 2018. Proceedings of the 14th International Conference on Statistical Analysis of Textual Data (Rome, 12-15 June 2018)*, a cura di Livia Celardo, Domenica Fioredistella Iezzi e Michelangelo Misuraca, 108-16. Roma: UniversItalia.

Farina, Annick, e Riccardo Billero. 2020. "Corpus in 'Natural' Language Versus 'Translation' Language: LBC Corpora, A Tool for Bilingual Lexicographic Writing." In *Text Analytics Advances and Challenges*, a cura di Domenica Fioredistella Iezzi, Damon Mayaffre e Michelangelo Misuraca, 167-78. Springer International.

Farina, Annick, e María Carlota Nicolás Martínez. 2020. "La banca dati LBC." In *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, a cura di Riccardo Billero, Annick Farina e María Carlota Nicolás Martínez, 7-17. Firenze: Firenze University Press.

Farina, Annick, e Lorella Sini. 2020. "Il Corpus LBC Francese." In *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, a cura di Riccardo Billero, Annick Farina e María Carlota Nicolás Martínez, 77-99. Firenze: Firenze University Press.

Friginal, Eric, e Jack Hardy. 2014. *Corpus-based Sociolinguistics. A guide for students*. London: Routledge.

Kilgarriff, Adam, Baisa Vít, Bušta Jan, et al. 2014. "The Sketch Engine: ten years on." *Lexicography* 1: 7-36.

Laviosa, Sara. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications.* Amsterdam-New York: Rodopi.

L'Homme, Marie-Claude. 2004. *La terminologie: principes et pratiques*. Montréal: Les Presses de l'Université de Montréal.

McEnery, Tony, e Andrew Hardy. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

GDT = Office Québécois de la Langue Française, *Grand Dictionnaire Terminologique* (GDT). https://gdt.oqlf.gouv.qc.ca (22/12/2022).

Palustre, Léon. 1892. *L'Architecture de la Renaissance*. Paris: Ancienne Maison Quantin, Librairies-Imprimeries réunies, May & Motteroz.

Perotto, Monica, e Valeria Zotti. 2022. "Il patrimonio culturale e artistico di Bologna nel corpus LBC (Lexikon of Cultural Heritage) e nel corpus UniCittà (UniCity)." In *Il patrimonio culturale della Biblioteca Universitaria di Bologna e della città allo specchio dei viaggiatori europei. Esplorazioni tra la prima modernità e l'era contemporanea*, a cura di Chiara Conterno e Fiammetta Sabba, 161-70. Bologna: Bologna University Press.

Williams, Geoffrey. 2005. "Introduction." In *La linguistique de corpus*, a cura di Geoffrey Williams, 13-18. Rennes: Presses universitaires de Rennes.

Williams, Geoffrey. 2020. "Architecture in the 1701 Dictionnaire Universel: Encoding and Analysing Architectural Terminology with Digital Humanities Methodologies." In *The Language of Art and Cultural Heritage: a Pluringual and Digital Perspective*, a cura di Valeria Zotti e Ana Pano Alamán, 190-207. Newcastle upon Tyne: Cambridge Scholars.

Zanettin, Federico. 1998. "Bilingual Comparable Corpora and the Training of Translators." *Meta: Journal des traducteurs / Translators' Journal* 43: 613-30.

Zanola, Maria Teresa. 2018. "La terminologie des arts et métiers entre production et commercialisation : une approche diachronique." *Terminàlia* 17: 16-23. http://terminalia.iec.cat (22/12/2022).

Zotti, Valeria. 2017. "L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle *Vite* di Vasari." In *Informatica umanistica: risorse e strumenti per lo studio del lessico dei beni culturali*, a cura di Valeria Zotti e Ana Pano Alamán, 105-33. Firenze: Firenze University Press.

Zotti, Valeria. 2021. "Les termes des arts dans les dictionnaires de la tradition française et dans les corpus de dernière génération : une relation d'inclusion réciproque?" In *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7-9 September 2021, Alexandroupolis*, vol. 2, a cura di Zoe Gavriilidou, Lydia Mitits e Spyros Kiosses, 683-91. Alexandroupolis: Democritus University of Thrace.

Zotti, Valeria. 2023. "La langue des Beaux-Arts : dimension interculturelle et enjeux terminologiques des corpus comparables." In *Terminologia e interculturalità. Problematiche e prospettive*, a cura di Maria Teresa Zanola e Paola Puccini, 229-245. Città di Castello: I libri di Emil.

Zweigenbaum, Pierre, e Benoît Habert. 2006. "Les corpus naissent tous comparables en droit : apports méthodologiques de l'acquisition lexicale en contexte multilingue." *Glottopol* 8: 22-24.

Sources citées du corpus BER

Buchot, Emmanuel. 2017. *Arcades de Bologne en Italie*, in *Blog carnet photographique – voyages – photos – infos* www.voyagesphotosmanu.com/arcades-de-bologne.html (22/12/2022).

Cochin, Charles-Nicolas. 1758. "Bologne." In *Le voyage d'Italie de Charles-Nicolas Cochin (1758)*, édité en fac-similé avec une introduction et des notes par Christian Michel, Publications de l'École française de Rome, 145. [1991]: 253-95.

Cotteau, Gustave. 1889. "Chapitre V Bologne - II." In *Le préhistorique en Europe : congrès, musées, excursions*, 67-88. Paris: J.-B. Bailliere.

Couture, Jean-Baptiste, 1926. "Chapitre XV." In *En Europe Notes de Voyage*, 74-81. Maine: Lewiston.

De Staël, Madame de. 1807. "Chapitre VII." In *Corinne, ou L'Italie*, t. II, l. XIX, 460-64. Paris: H. Nicolle Librerie Stereotipe.

Huchet, Bernard. 2011. "La gravure à Bologne : entre sacré et profane, 1560-1660." *Nouvelles de l'Estampe* 233-34: 58-60.

Jestaz, Bertrand. 1966. *Le Voyage d'Italie de Robert de Cotte. Étude, édition et catalogue des dessins*. Roma: École française de Rome.

La Lande, Jérôme de. 1769. "Chapitre VI : Des Eglises qui sont hors de la Ville." In *Voyage d'un françois en Italie, fait dans les années 1765 et 1766, contenant l'histoire, les anecdotes les plus singulières de l'Italie, & sa description, les mœurs, les usages*, t. 2, 99-110. Paris: Venise.

La Porte, Joseph de. 1779. “Lettre CCCXXVI : Suite de l’Etat ecclésiastique.” In *Le voyageur françois, ou la Connoissance de l’ancien et du nouveau monde, mise au jour par M. l’Abbé Delaporte*, t. XXVI, 29-64. Paris: L. Cellot.

La Roche, Jean de. 1783. *Voyage d’un amateur des arts, en Flandre, dans les Pays-Bas, en Hollande, en France, en Savoye, en Italie, en Suisse, fait dans les années 1775-76-77-78*, t. III, 140-203, Amsterdam: Desoer.

Marmottan, Paul. 1918. “Napoléon à Bologne (21 au 24 Juin 1805).” *Revue des Études Napoléoniennes: Les Origines de l’Europe Nouvelle*, 249-62. Paris: F. Alcan.

Plessis, Joseph-Octave. 1903. *Journal d’un voyage en Europe*. Québec: Librairie Montmorency-Laval, Pruneau & Kirouac.

Rémusat, Charles de. 1861. “L’Italie, Notes de Voyage - Deuxième partie - I. Bologne.” *Revue des Deux Mondes* 35: 335-49.

Robida, Albert. 1878. *Les vieilles villes d’Italie: notes et souvenirs*. Paris: Maurice Dreyfous Editeur.

Rocheleau, Eustache. 1922. *XVI. BOLOGNE*, in *Journal d’un pèlerin : Trois-Rivières, Assise, Rome*, 113-121. Québec: s.n.

s.a. “Bologne et ses tours attirent les touristes.” *La Presse*, Montréal, 9 novembre 1968. 262: 43.

Starke, Richard, e Mariana Starke. 1833-1834. “De Milan à Bologne.” In *Guide du voyageur en Italie*, 161-64. Paris: Audin.

Stendhal. 1932. *Écoles italiennes de peinture*. Paris: Le Divan.

Taillasson, Jean-Joseph. 1807. *Observations sur quelques grands peintres*. Duminil-Lesueur.

Viollet Le Duc Emmanuel-Louis-Nicolas. 1854-1868. *Dictionnaire raisonné de l’architecture française du XIe au XVIe siècle*. Paris: Bance editeur.